

MODELLING THE VALUE OF LOCATION IN THE PREDICTION OF RESIDENTIAL PROPERTY VALUE

by

Abdul Hamid bin Hj. Mar Iman

Centre for Real Estate Studies

e-mail: hamid@fksg.utm.my

Chin Chui Vui

and Department of Property Management

e-mail: chinchuivui@yahoo.com

*Faculty of Engineering & Geoinformation Science
Universiti Teknologi Malaysia*

Abstract

The important of location to effect property value is widely acknowledged. However, traditional approaches in quantifying location influence on property value modelling are unsatisfactory. Value Residual Surface (VRS) has been suggested as an alternative to resolving the difficulty in the traditional modelling of location influence on property values in a particular area. This paper discusses the use of GIS generated VRS for delineating residential location value factors in the multiple regression analysis (MRA) model. A sample of 125 single and double-storey residential properties was used to construct regression model. Residuals generated from location-blind model were used to construct VRS to develop location adjustment factors. The results showed that using VRS integrative has allowed a clearer spatial visual picture of the location influence to be captured at all level in the study area. And, the location factors influence to property value can be modelled in a more effective way.

Keywords: Geographical Information System (GIS), Location factors, Multiple Regression Analysis (MRA), Residential Property Values, Value Residual Surface (VRS)

1.0 INTRODUCTION

Residential is a multi-dimensional heterogeneous commodity, characterized by durability and structural inflexibility as well as spatial immobility. It has a unique bundle of attributes such as accessibility to work, transport, amenities, physical characteristics, neighbourhood, and environmental quality (Muth, 1960; Ridker and Henning, 1967; Stegman, 1969; Kain and Quigley, 1970; Evans, 1973; Lerman, 1979; So et al., 1997). Many of these attributes are spatially-related in the form of popularly known “location, location, location” hierarchy (Pearson, 1991). Real estate is spatially unique in which location is an intrinsic attribute that directly determines the quality and market value of the property.

However, modelling the location factors in property valuation has proved difficult because of the wide range of spatially definable attributes, which may or may not affect value at a particular time and location. Furthermore, there is little literature consensus as to the best proxy for location factor measurement.

Multiple Regression Analysis (MRA) is considered a classical and primary technique for explaining and predicting property values whereby location factors can potentially be taken into account. In particular, MRA has been used to estimate residential property values in the U.S since the 1950s and in the U.K since the 1980s (Pendleton, 1965; Greaves, 1984; Adair and McGreal, 1988). It was also applied in other countries such as Australia, New Zealand, and Singapore, but has yet to be adopted in Malaysia.

In applying MRA, valuers identify the data to be specified and measured in quantitative form. This task becomes more complicated when the location influence on property values need to be explicitly identified and modelled.

This article discusses the use of value residual surface (VRS) generated using the combination of Geographical Information System (GIS) and Multiple Regression Analysis (MRA) in a “hybrid” predictive model that utilizes VRS to create a location factors adjustment.

The second part of this paper briefly reviews the location modelling, followed by a brief description of the study area is discussed in the third section. Data and analysis procedure are discussed in the fourth section. Section of results and discussion follows then. The final part of this paper concludes the study.

2.0 THE MODELLING OF LOCATION

Location is an amalgam of several factors that includes a number of spatial elements such as accessibility to shopping, employment, educational and leisure facilities; exposure to adverse environmental effects such as traffic noise and hazard; neighbourhood amenity; perceived levels of neighbourhood security; and so on (Gallimore et al., 1996, McCluskey et al., 2000). Generally, these can be divided into

two, neighbourhood quality and accessibility (McCluskey et al., 2000). Few, however, are capable of numeric measurement but, even that, the measures may not always be valid representation of the influence, especially because of the complex interaction of value factors. For example, the common approach to examining location influence on property values is to include a distance variable from the central business district (CBD), assuming homocentric locations. This is based on the traditional location theory that examines the role of accessibility to central locations on property prices. There are also researchers who employ more sophisticated measurements in measuring location such as using the type of transport, time taken per trip, and transportation cost. However, house prices are determined not only by accessibility but also by the environmental attributes of location (Stegman, 1969; Richardson, 1971; and Pollakowski, 1982).

There are also theories of multiple-nuclei model incorporating the concentric pattern that are more appropriate for analyzing location influence on property values. For example, pattern of property values may reflect the influence of satellite towns rather than that of regional centres (see Hamid, 2001). On the basis of these theories, location partitioning is analyzed by using location dummy variables (Hamid, 2003). Essentially, this is to subdivide a particular geographic area into “realistic” sub-market or neighbourhood. However, this could pose modelling constraint in terms of data representativeness when some neighbourhoods with too few transactions give rise to “small sample” problems in the statistical estimation. Besides, there is little consensus, however, on which variables are the best proxy for neighbourhood quality measurement. When neighbourhood is applied, a hard edge may be implied at such boundaries, whereas in reality the varying influence of location may operate far more smoothly and the spatial trends occur as opposed to distinct areas of homogeneous property subsets (Gallimore et al., 1996 and Mackmin, 1989).

Given the limitation and highly complex process for discrete measurement of location have encouraged researchers to search for alternative approaches to derive location compensation factors. The value residual surface (VRS) was suggested with the emergence of GIS technology as an alternative to resolving the difficulty in the traditional modelling of location influence on property values in a particular area.

In constructing the value of location through this approach, the discrepancies between the actual and location-blind model are transformed into a 2½-dimensional value residual surface (VRS) using GIS. The latitude and longitude (x-y) co-ordinates indicate the geo-referenced location of each particular property, while the generated residuals are depicted on the vertical axis (z-axis). This residual surface could then be used to adjust for the under or over valuation of any property within the area to estimate a location influence. This variable is then included, along with other variables, in a multiple regression or other model to capture location influences.

3.0 ANALYSIS PROCEDURE

The study area selected comprised three adjoining residential areas in the sub-urban of Johor Bahru city. They were Taman Pelangi, Taman Sentosa, and Taman Sri Tebrau, selected particularly due to their active sales transaction. The residential sales data were obtained from the Johor Bahru Valuation and Property Services Office (JPPH) covering a period of 3 years from January 2001 to December 2003. This was a period during which local advisers felt that the movement of property prices in the locality was quite stable. The sales records included sales price, transacted date, lot number, property characteristics, and property address. Property addresses were particularly recorded to ensure that the location of each property could subsequently be digitized with acceptable degree of accuracy.

A total of 125 terraced residential units in the three adjoining housing scheme were used for estimating the models. The distribution of the sales data in the study area is shown in Figure 2.

In order to calculate spatial residuals, location-blind modelling was performed by including only data which related to the physical characteristics of the properties. Also, due to data limitation, the variables used in this study just represent some property characteristics only. Therefore, some form of model misspecification can be expected. The *Statistical Package for the Social Science* was used the regression runs. The variables included and the units of measurement in the location-blind model are shown as below:

Land Area (LA) - Square meters (sq. m.)

Gross Floor Area (GFA) - Square meters (sq. m.)

Ancillary Area (AA) - Square meters (sq. m.)

Type of Terrace (T1) - Dummy (single story terrace = 1)

Age (AGE) - Years

Condition of the Building (GCOND) - Dummy (good in condition = 1)

Floor Finishes (FFNISH2) – Dummy (class 2 - combination of parquet, terrazzo, semen, mosaic or tiles floor finishes = 1)

No. of Bedroom (3BEDR) – Dummy (3 bedrooms = 1)

Box-Cox test was applied to determine the best functional forms to be chosen. Besides, diagnostic tests for multicollinearity, autocorrelation, heteroscedasticity, and model misspecification were performed.

For location value adjustment, the residuals from the location-blind regression model were used to construct VRS and these residuals were expressed in percentage terms. The VRS was generated by using GIS. In building VRS, the surface was generated by calculating weights according to the reciprocal of the distance between the subject and the neighbouring residuals. Residuals within a 500-meter radius from the subject property were being identified.

The models were evaluated on the basis of comparative adjusted R^2 , F-value, SEE, and SSE, and “change of significance” of a particular variable specified.

4.0 RESULTS AND DISCUSSION

Table 1 shows the basic descriptive statistics of the sample. About 48% of the sample size was made up of properties in Taman Pelangi, followed by Taman Sentosa (35%), and Taman Sri Tebrau (17%).

The mean land area of the properties was about 164 sq. m. The range of land area was purposely controlled in the sample to avoid too much variance of this variable. On average, the ratio of gross floor area to land area was approximately 3.7:5.0. The mean ancillary area of the properties was about 32 sq. m.

On the basis of standard deviation, gross floor area has more variation compared to land area and ancillary area. This means, the people in the market could have been more concerned about differences in floor area, making this factor a possibly more influential physical characteristic of residential properties.

The sample was rather dominated by double-storey 4-bedroom terraced residential. Besides, the majority of the properties, single- and double-storey terraced alike, were of good building condition and have either parquet, terrazzo, cement, mosaic or tile floor finishes. The mean age of the residential properties was about 22 years. Since these are old housing schemes, the majority of the sampled residential properties have also undergone some renovation.

Table 1: Descriptive statistics of the sample variables

	Minimum	Maximum	Mean	Std. Deviation
Price (RM)	160,000.00	445,000.00	263,125.61	60,429.89
Land Area (sq.m.)	141.02	216.82	163.67	14.80
Gross Floor Area (sq.m.)	69.21	218.32	121.78	34.50
Ancillary Area (sq.m.)	7.43	91.14	31.78	12.37
Single Storey Terrace	0.00	1.00	0.46	0.50
Double Storey Terrace	0.00	1.00	0.54	0.50
Age of Building (years)	15.00	29.00	22.41	3.21
Good Building Condition	0.00	1.00	0.99	0.09
Unsatisfactory Building Condition	0.00	1.00	0.008	0.09
Class 2 ¹ Floor Finishes	0.00	1.00	0.97	0.18
Class 1 ² Floor Finishes	0.00	1.00	0.03	0.18
3 Bedrooms	0.00	1.00	0.41	0.49
4 Bedrooms	0.00	1.00	0.59	0.49
Taman Pelangi	0.00	1.00	0.48	0.50
Taman Sentosa	0.00	1.00	0.35	0.48
Taman Sri Tebrau	0.00	1.00	0.17	0.38

1 – combination of granite, marble, or homogeneous floor finishes.

2 - combination of parquet, terrazzo, cement, mosaic or tile floor finishes.

Based on Box-Cox transformation, the log-log function was the most appropriate model to choose. For this reason, only the log-log function is reported in this study.

Model I in Table 2 shows the results of the location-blind model. This model was used as a basis for generating residuals for prediction surface later. This model explained about 83 percent variation in the prices of residential property in this study. Land area, gross floor area, age of the building, and the building condition were statistically

significant in this model. Their coefficient signs show that all of these variables were theoretically plausible.

This adjustment factor can be created and included into the original location-blind model utilizing information from the VRS surface. Using VRS (Figure 1 and Figure 3), the location price adjustment component was created in the second-stage model. This location component was represented by the variable `LOCATION` in the hybrid model – Model II (Table 2).

The diagnostic tests indicated no evidence of multicollinearity, heteroscedasticity and autocorrelation in the hybrid model. Breusch-Pagan test showed that the model has Q-values smaller than critical value of 2.730. The Durbin Watson test also rejected the hypothesis that the model has negative or positive autocorrelation. However, the RESET has indicated model misspecification. The main reason for this phenomenon could be the exclusion of some other important physical variables of the residential properties such as architectural style, car porch, number of toilets, etc in the model. Because of this, it could be expected that the residential price estimates of some residential attributes included in the model could have been biased. Again, the interpretation of the regression coefficients should be made with some caveat.

Table 2 shows the statistical performance of the model that uses GIS-MRA-generated VRS (Model II). When the residuals from the VRS were putted into the model as location value, the results show increases in explanation of the model with Adjusted R^2 and R^2 99.9 percent respectively. In terms of F-value, this model (Model II) also has a higher level of significance. Further, Model II has lower standard error of estimate (SEE) and sum squared errors (SSE). Besides, overall, the t-values for all regressors in Model II were changed from being insignificant to being significant and achieved higher t-values. These figures indicated that residuals in the model were capturing most of the location effect.

Table 2: Comparison of the preliminary and hybrid model

Model	Model I	Model I
Model form	Double log	Double log
R ²	0.828	0.999
Adjusted R ²	0.816	0.999
F-value	69.944	12284.562
SEE	4.285E-02	3.348E-03
SSE	0.213	1.289E-03
Dependent variable (LG_PRIC)	House price (LG_PRIC)	House price (LG_PRIC)
Independent variables:	Coefficient	Coefficient
(Constant)	3.761	3.767
	(12.945)***	(165.955)***
Log of Land Area (LG_LA)	0.342	0.345
	(2.916)***	(37.598)***
Log of Gross Floor Area (LG_GFA)	0.528	0.524
	(7.165)***	(91.051)***
Log of Ancillary Area (LG_AA)	0.034	0.036
	(1.279)	(17.199)***
Type of Terrace (Single storey) (T1)	-0.023	-0.022
	(-1.292)	(-15.945)***
Log of Age (LG_AGE)	-0.277	-0.282
	(-4.194)***	(-54.666)***
Condition of the Building (Good) (GCOND)	0.142	0.147
	(3.174)***	(42.217)***
Floor Finishes (Class 2) (FFNISH2)	-0.014	-0.016
	(-0.611)	(-9.416)***
No. of Bedroom (3 bedrooms) (3BEDR)	0.008	0.007
	(0.622)	(6.844)***
Location		0.004
		(137.439)***

***Significant at .01. Model I is the preliminary model without location factor. Model II used VRS as location factors.

A two-dimensional view of VRS is shown in Figures 1. Similarly, Figures 3 shows the VRS in a 2½-dimensional view. These visual profiles provide an overall idea about at least two things. First, they force or perhaps more appropriately “help” valuers to find explanation for location factors that give rise to “bumps” and “potholes” or “ridges” and “trenches” (Figure 3) in the study area. This necessarily needs the knowledge and familiarity of the local surroundings and property market.

From site inspection, it was observed that the “bumpy” areas represent neighbourhoods that were quite strongly influenced by positive location elements such as proximity to shopping centres, shop houses, leisure facilities, and open space or park (see Figure 3).

The appeal of business activities, neighbourhood, and environmental qualities in these areas has tended to bid up the sale prices of residential properties nearby. Examples of such areas can be found to the southeast of Taman Sri Tebrau and, to the middle and northeast of Taman Pelangi.

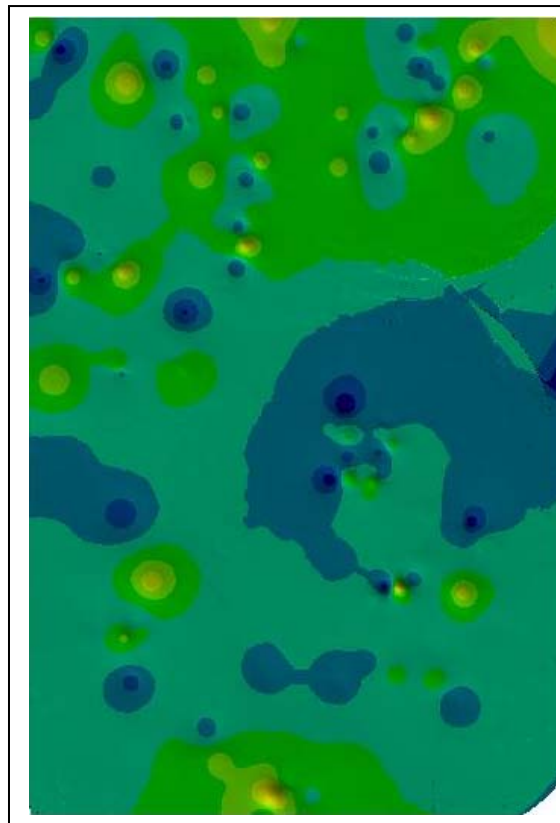


Figure 1: 2-D view of VRS

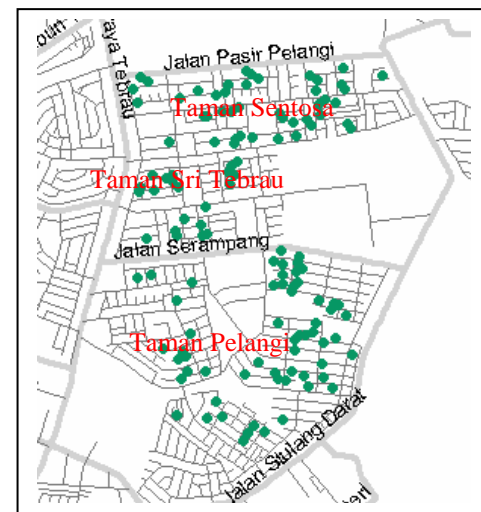
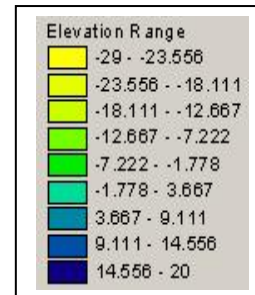


Figure 2: Sample distribution

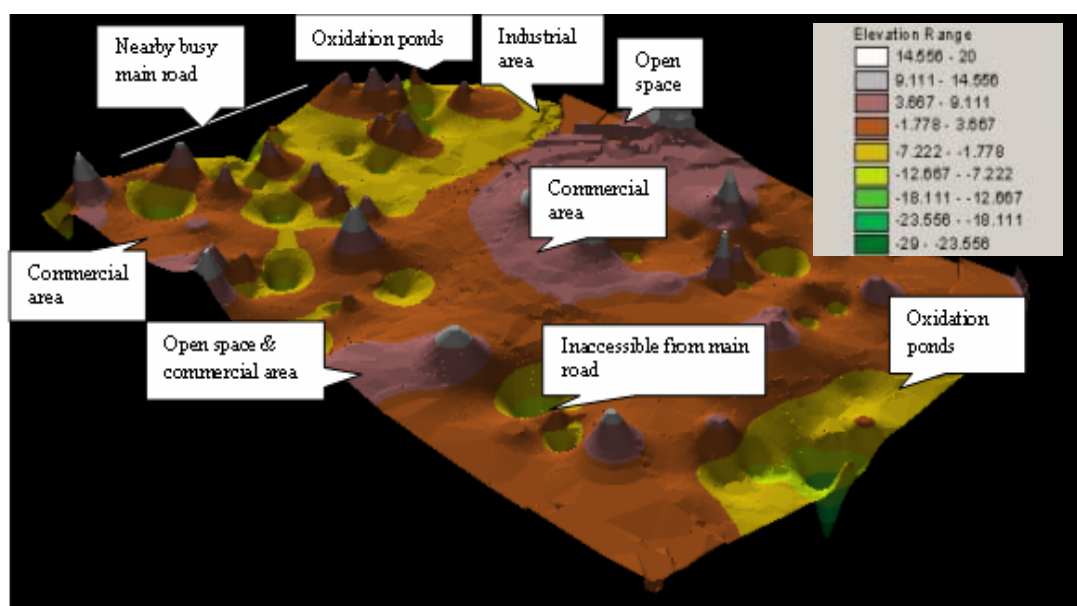


Figure 3: 2½-D view of VRS

On the other hand, being located to the east of Taman Sentosa and to the south of Taman Pelangi, the proximity of houses to negative location elements such as oxidation ponds, industrial area, and noisy main roads was seen to be associated with “potholes” or “trenches”. Besides, the structural designs of residential units in these “trenchy” neighbourhoods were not considered superior than those in the “ridgey” neighbourhoods. These houses in these areas were also far away from any shop houses and other commercial activities.

Second, how location adjustment should be made across the study area. The “bumps” or “ridges” represent under-valued while “potholes” or “trenches” represent over-valued residential properties. Consequently, the “bumpy” or “ridgey” areas need upward while the “potholey” or “trenchy” areas need downward location price adjustment.

5.0 CONCLUSION

The primary objective of this study was to examine the usefulness VRS in creating location value adjustment factors in the prediction of property values by taking residential properties as a study case. The generated surface has enabled a more visualized representation of location influence on property values and has enabled such influence to be captured at any geographic points across a particular area. The GIS-MRA-generated VRS has managed to improve overall model’s statistical quality.

The results from this study should be able to encourage further research in various aspects of spatial modelling in property valuation. Future research may focus on the refinement of GIS-MRA generated VRS techniques using a more complete model specification. The effect of sample structure on the resulting VRS can also be examined further. Sample structure includes elements like the number, distribution, and type (single- or double-storey property value, per square foot or per unit property value) of observation points.

Further investigation on the likely spatial elements that give rise to “bumps” or “ridges” and “potholes” or “trenches” can be further explored in future studies. By doing this, some factors can be statistically explained with certain location detail in the local

market. As a matter of fact, the practice in the property valuation is that people naturally want to discover the specific location factors that have actually influenced property values in the local market. This will allow back-to-back approaches between GIS-MRA VRS techniques and the traditional MRA to be adopted from time to time.

REFERENCES

- Adair, A. S. and McGreal, W.S. (1988). The Application of Multiple Regression Analysis in Property Valuation. *Journal of Valuation*. 6(1): 57-67.
- Evans, A.W. (1973). *The Economics of Residential Location*. London: Macmillan.
- Gallimore, P., Fletcher, M. and Carter, M. (1996). Modelling the Influence of Location on Value. *Journal of Property Valuation and Investment*. 14(1): 6-19.
- Greaves, M. (1984). The Determinants of Residential Values: The Hierarchical and Statistical Approaches. *Journal of Valuation*. 3: 5-23.
- Hamid, Abdul, bin Mar Iman (2001). *Incorporating Geographic Information System in the Prediction of Farm Property Values*. Unpublished Ph.D thesis. Lincoln University, New Zealand.
- Hamid, Abdul, bin Mar Iman (2003). Priced-Contour Based Spatial Dummy Variables for Segmenting Market in the GIS-Hedonic Modelling of Residential Property Prices. *International Conference on Urban Development and Management*, Langkawi, Malaysia, 7-9 July.
- Kain, J. and Quigley, J. (1970). Measuring the Value of Housing Quality. *Journal of the American Statistical Association*. 45: 532-548.
- Mackmin, D. (1989). *The Valuation and Sale of Residential Property*. London: Routledge.
- McCluskey, W. J., Deddis, W. G., Lamont, I. G., and Borst, R. A. (2000). The Application of Surface Generated Interpolation Models for the Prediction of Residential Property Values. *Journal of Property Investment and Finance*. 18(2): 162-176.
- Muth, R. F. (1969). *Cities and Housing*. Chicago, II: University of Chicago Press.
- Pearson, T.D. (1991). Location! Location! Location! What is Location!. *The Appraisal Journal*, LIX(1): 7-20.
- Pendleton, E. C. (1965). Statistical Inference in Appraisal and Assessment Procedures. *The Appraisal Journal*. 37: 501-512.
- Pollakowski, H. O. (1982). *Urban Housing Markets and Residential Location*. D.C. Lexington, MA: Health and Company.

Richardson, H. W. (1971). *Urban Economics*. Penguin: Harmondsworth.

Ridker, R. and Henning, J. (1967). The Determinants of Residential Property Values with Special Reference to Air Pollution. *Review of Economics and Statistics*. 49: 246-257.

So, H. M., Tse, R.Y. C. and Ganesan, S. (1997). Estimation the Influence of Transport on House Prices: Evidence from Hong Kong. *Journal of Property Valuation & Investment*. 15(1): 40-47.

Stegman, M. A. (1969). Accessibility Models and Residential Location. *Journal of American Institute of Planners*. 35: 22-29.